



**You have downloaded a document from
RE-BUS
repository of the University of Silesia in Katowice**

Title: Językoznawstwo korpusowe – próba oceny nowych narzędzi badawczych na podstawie książki Corpus Linguistic for Grammar. A Guide for Research Christiana Jonesa i Daniela Wallera (London–New York 2015, 201 s.)

Author: Kinga Wąsińska

Citation style: Wąsińska Kinga. (2019). Językoznawstwo korpusowe – próba oceny nowych narzędzi badawczych na podstawie książki Corpus Linguistic for Grammar. A Guide for Research Christiana Jonesa i Daniela Wallera (London–New York 2015, 201 s.). "Forum Lingwistyczne" (nr 6 (2019), s. 145-150), doi 10.31261/FL.2019.06.11



Uznanie autorstwa - Na tych samych warunkach - Licencja ta pozwala na kopiowanie, zmienianie, rozprowadzanie, przedstawianie i wykonywanie utworu tak długo, jak tylko na utwory zależne będzie udzielana taka sama licencja.



UNIwersYTET ŚLĄSKI
W KATOWICACH



Biblioteka
Uniwersytetu Śląskiego



Ministerstwo Nauki
i Szkolnictwa Wyższego

Językoznawstwo korpusowe –
próba oceny nowych narzędzi badawczych
na podstawie książki *Corpus Linguistic for Grammar.*
A Guide for Research Christiana Jonesa i Daniela Wallera
(London–New York 2015, 201 s.)

Corpus Linguistics – an Attempt to Evaluate New Research
Tools Based on a Book *Corpus Linguistic for*
Grammar. A Guide for Research by Christian Jones and
Daniel Waller (London–New York 2015, 201 pp.)

Gromadzenie materiału językowego do badań naukowych kojarzone dotychczas ze żmudnym czytaniem tekstów i tworzeniem fiszkowych katalogów powoli odchodzi do przeszłości. Rozwój technologii informacyjno-komunikacyjnych sprawił, że obcowanie z tekstem na przestrzeni ostatnich lat znacznie się zmieniło, a wraz z nim – sposoby badania zdigitalizowanych dokumentów. Dużym ułatwieniem dla badaczy było pojawienie się w ogólnodostępnych bibliotekach cyfrowych¹ dokumentów w formacie DjVu, choć prawdziwym przełomem nazwać można dopiero powstanie repozytoriów z narzędziami do przeszukiwania zgromadzonych danych. Taką nową jakość pracy przyniosły internetowe korpusy językowe, które dostarczają informacji na temat typowych użycí słów i konstrukcji oraz na temat ich znaczenia i funkcji. Jak piszą twórcy Narodowego Korpusu Języka Polskiego – największego lingwistycznie anotowanego korpusu współczesnej polszczyzny: „Bez dostępu do korpusu nie da się dziś prowadzić badań językoznawczych, pisać słowników ani podręczników języków obcych, tworzyć wyszukiwarek uwzględniających polską odmianę, tłumaczyć komputerowych ani innych programów zaawansowanej technologii językowej” (NKJP: online).

Widoczny jest wzrost zainteresowania badaniami korpusowymi. Choć publikacji powstających dzięki wykorzystaniu danych korpusowych przybywa, to ciągle brakuje polskich

* e-mail: kinga.wasinska@us.edu.pl

¹ W Polsce pierwszą biblioteką cyfrową była Polska Biblioteka Internetowa. Wstępne założenia projektu tej biblioteki opisano w Programie Powszechnej Informatyzacji Nowoczesna Polska (przyjętym jako program przedwyborczy przez SLD 8 kwietnia 2001 r.), a oficjalnie została ona udostępniona 1 września 2003 r. Przedsięwzięcie to ocenia się jednak jako mało udane ze względu na powolny przyrost zdigitalizowanych zasobów. W październiku 2002 r. uruchomiono Wielkopolską Bibliotekę Cyfrową. Jej dynamiczny rozwój zapoczątkował w Polsce proces tworzenia bibliotek cyfrowych (GŁOWACKA, 2011).

opracowań i podręczników charakteryzujących tę badawczą metodę pracy². Literatura obcojęzyczna wypełnia powstałą lukę z nadstatkiem. Na temat korpusów pisze się w Europie i na świecie dużo, w formie zarówno artykułów, jak i monografii³. Bogato reprezentowana jest literatura korpusowa o nachyleniu glottodydaktycznym. W związku z tym zdecydowałam się przybliżyć książkę *Corpus Linguistic for Grammar. A Guide for Research*, która ukazała się w 2015 roku nakładem wydawnictwa Routledge, wydającego serię *Corpus Linguistics Guides*. W serii tej ukazały się dotychczas między innymi: *Corpus Linguistics and the Analysis for Sociolinguistic Change* (Joan O'SULLIVAN, 2019), *Corpus Linguistics for Online Communication* (Luke COLLINS, 2019), *Corpus Linguistics for Pragmatics* (Christoph RÜHLEMANN, 2018), *Corpus Linguistics for Discourse Analysis* (Michael HANDFORD, 2018), *Corpus Linguistics for ELT* (Ivon TIMMIS, 2015).

Wybór tej właśnie pozycji, kierowanej głównie do nauczycieli języka angielskiego jako obcego, podyktowany był pewną ciekawością poznawczą. Językoznawcy interesujący się historycznym rozwojem języka polskiego – do grona których należą – często podnoszą kwestię posiadania niepełnej kompetencji językowej potrzebnej do badań dawnej polszczyzny. Zwracają też uwagę na nietatwe obcowanie z historycznym materiałem językowym. Zainspirowana tymi trudnościami postanowiłam poznać tematykę badań korpusowych glottodydaktyków, ich metody pracy i problemy, z którymi się mierzą. Zarówno lektorom, jak i historykom języka przyświeca przecież jeden cel: jak najpełniejsze poznanie języka.

Przypomnijmy, że językoznawstwo korpusowe to część językoznawstwa komputerowego, dział na styku informatyki i językoznawstwa. Na początku kojarzone było wyłącznie z gromadzeniem materiałów językowych w formie elektronicznej. Z czasem, gdy możliwe stało się przeszukiwanie zdigitalizowanego tekstu z użyciem wyspecjalizowanych narzędzi, z metody marginalnej przekształciło się w pożądaną i cenioną sposób pracy. W związku z tym, że pierwsze korpusy dotyczyły języka mówionego, najdłuższą historię stosowania metody korpusowej mają badania zorientowane kognitywnie. Obecnie z wielorakich typów korpusów korzystają badacze różnych nurtów dyscyplin humanistyki i nauk społecznych (PRZEPIÓRKOWSKI i in., 2009; PRZEPIÓRKOWSKI i in., red., 2012).

Autorzy książki *Corpus Linguistic for Grammar* – Christian Jones i Daniel Waller – są nauczycielami języka angielskiego dla obcokrajowców z ponad dwudziestoletnim stażem. Jones nauczał angielskiego w Japonii i Tajlandii, z kolei Waller wykładał głównie w Wielkiej Brytanii i Turcji. Z różnych doświadczeń glottodydaktycznych autorów wyrosła potrzeba napisania podręcznika na temat językoznawstwa korpusowego, który – dzięki bogactwu tychże doświadczeń – jest pozycją praktyczną, zawierającą wiele sprawdzonych dydaktycznie ćwiczeń.

Omawiana książka podzielona została na trzy części; na każdą z nich składają się po trzy rozdziały. W części pierwszej (*Defining Grammar and Using Corpora*) zamieszczone zostały definicje i wyjaśnienia terminologiczne. Ustalenia te są dość ważne – zwłaszcza dla nauczycieli uczących języka obcego, choć nie tylko. Uświadamiają, że wyniki badań

² Można wskazać jedynie pojedyncze prace: LEWANDOWSKA-TOMASZCZYK, red., 2005; PRZEPIÓRKOWSKI i in., red., 2012.

³ Jako przykłady ważniejszych pozycji książkowych można wymienić: SINCLAIR, 1991; BIBER, CONRAD, REPPEN, 2008; LÜDELING, KYTÖ, 2010; MCENERY, HARDIE, 2011.

mogą różnić się w zależności od sposobu podejścia do kluczowych zagadnień, między innymi od tego, co rozumiemy pod pojęciem poprawnie (gramatycznie) zbudowanego zdania. Pouczający może być fakt, że skuteczna komunikacja nie wymaga posługiwania się takimi zdaniami. Warto zatem obserwować struktury językowe przynależne do niższego rejestru języka, a także realizacje gramatyczne charakterystyczne dla uzusu językowego.

Część druga książki, zatytułowana *Corpus Linguistics for Grammar: Areas of Investigation*, zawiera wskazania konkretnych wartości, które można wyszukiwać za pomocą narzędzi dostępnych w korpusach językowych. Część tę podzielono na trzy rozdziały zgodnie z trzema zakresami tematycznymi badań korpusowych. Jako pierwszą wartość wyszukiwania wskazano częstotliwość występowania synonimicznych struktur gramatycznych w tekście (*Frequency*). Kolejne dwa rozdziały tej części⁴ zawierają obszernie omówienie sposobów wykorzystywania korpusów do badań nad elementami języka, których znaczenie aktualizuje się w zapisie mowy: czy to przez zastosowanie szablonowych struktur zdaniowych przywoływanych z pamięci zamiast tworzenia każdorazowo nowych zdań, czy poprzez zabarwienie emocjonalne językowo wyrażane szykiem wyrazów w wypowiedzi. Rozdziały te zatytułowano: *Chunks*⁵ and *Colligations* oraz *Semantic Prosody*. Optyka tej części książki wywodzi się z językoznawstwa kognitywnego, którego główna teza głosi, że leksyka powinna być rozpatrywana tylko w kontekście, a prowadzenie badań semantycznych możliwe jest jedynie z uwzględnieniem związków składniowych wyrazów w zdaniu.

Każdy z rozdziałów dwóch pierwszych części książki wypełniają różnego rodzaju ćwiczenia. Niektóre z zadań autorzy wykonują niejako razem z czytelnikiem – instruując go krok po kroku (*Bringing the Analysis Together*). Są też ćwiczenia w formie pytań otwartych do przeprowadzonych analiz lub poleceń uszczegóławiających to, co zostało już przekazane. Za każdym razem, kiedy czytelnik zauważa etykietę *Try It Yourself*, może rozważyć podsunięte przez autorów zagadnienia, po czym sprawdzić swoje pomysły, porównując wyniki własnych badań z dołączonymi na końcu podręcznika sugerowanymi odpowiedziami (*Suggested Answers*). Najbardziej rozbudowane i wielowątkowe polecenia zamieszczono w działach *Further Practice*. Te ćwiczenia wymagają konkretnej pracy na wskazanym korpusie, przeprowadzonej w całkowicie samodzielny sposób, co pozwala odbiorcy najpełniej sprawdzić swoje rozeznanie w metodzie badań korpusowych.

Przykładowo, jedno z ćwiczeń w punkcie *Further Practice* podsumowującym część pierwszą książki opatrzone zostało poleceniem:

Wykonaj podstawowe badanie korpusowe, przeszukując jeden z podanych w książce korpusów o wolnym dostępie w celu zgromadzenia wszystkich kontekstów, w których pojawia się wybrane przez ciebie słowo, fraza lub sentencja. Przyjrzyj się, czego możesz dowiedzieć się na podstawie otrzymanych danych. A następnie

⁴ Są to rozdziały 5. i 6., ponieważ numeracja kolejnych partii w książce jest ciągła i nie uwzględnia podziału na części.

⁵ Z uwagi na brak w polskiej terminologii odpowiednika terminu *chunks* przytaczam definicję autorów przyjętą za WRAY'EM (2005: 9): „A formulaic sequences is a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar”.

zastanów się, jakie dodatkowe dane byłyby ci potrzebne, abyś mógł pogłębić swoje analizy⁶ [tłum. K.W.].

s. 59

To polecenie, podobnie jak większość zadań w książce, kierowane jest do osób, które wcześniej nie miały częstego kontaktu z językoznawstwem korpusowym. Autorzy, dostarczając zarówno narzędzi do badań, jak i stosownych przykładów, chcą zmotywować czytelników do przeprowadzenia własnego dochodzenia naukowego, w którym dane pozyskane z korpusu byłyby bazą wyjściową, nie zaś zbiorem kontekstów wystarczającym do budowania na ich podstawie teorii i twierdzeń. Taka forma ćwiczeń pomaga również ocenić, czy umiemy wykorzystać atuty tej metody w interesujących nas zakresach poznawczych.

Oprócz wyszczególnionych działów ćwiczeniowych (w różnych partiach książki) w rozdziałach teoretycznych wprowadzone są *Sample Exercises*, które pozwalają czytelnikowi sprawdzić, czy w pełni rozumiał on dotychczas wyłożone treści.

Ostatnia część książki, zatytułowana *Applications of Research*, adresowana jest do trzech różnych grup odbiorców. Rozdział otwierający tę część⁷ napisany został głównie z myślą o nauczycielach uczących języka angielskiego jako języka pierwszego. Badacze zachęcają, by świadomie wybierać konteksty do nauki gramatyki, słownictwa i wymowy oraz koncentrować się na żywej, autentycznej mowie, której źródło stanowią właśnie korpusy tekstów. Co cenne, omawiają również ograniczenia użyteczności metody korpusowej w tym zakresie. Kolejny rozdział tej części⁸ (*Wider Applications: Data-driven Journalism and Discourse Analysis*) dotyczy badań dyskursologicznych. Opierając się na fenomenie poczytności publikacji internetowych, autorzy analizują styl dziennikarski. W ich ujęciu wszelkie groźby wtórnego analfabetyzmu i prognozy zaniku książki jako takiej rozwiązał rozwój elektronicznych nośników danych, które uatrakcyjniły kontakt z językiem pisanym. Jak dowodzą badacze, z Internetu jako źródła informacji coraz częściej korzystają dziennikarze i analitycy. Rozdział ten pokazuje więc, w jaki sposób odbywa się ten – w pewien sposób odwrotny – kierunek pozyskiwania wiedzy i jaki wpływ ma to zjawisko na kształt stylu nazywanego dziennikarskim. Ostatni rozdział książki⁹ kończą przykłady ściśle językowych projektów badawczych. Autorzy szerzej opisują temat badań kolokacji i zmian w obrębie łączliwości wyrazów – jeden z głównych atutów korpusowych analiz języka, który może stać się przyczynkiem do wysnucia wniosków na temat natury ewolucji znaczeń wyrazów. Omawiają również prace przeprowadzone z użyciem różnych korpusów języka angielskiego o wolnym dostępie, między innymi BYU-BNC, GloWbe, WebCorpLSE, COCA, COBUILD. Analizowany w książce przykład obserwacji znaczenia wyrazu *bloody* 'krwawy' (w zależności od rodzaju/stylu czasopisma, w którym to słowo występuje) bezpośrednio można odnieść do języka polskiego – na przykład w zakresie użycia słów typu *cholernie* (czuć się *cholernie* 'czuć się kiepsko, marnie' *vs.* *cholernie* *zdolny* 'bardzo zdolny') czy *straszenie*

⁶ „Do a basic search for a form in any of the corpora listed above. This could be at the word, clause or sentence level. What does the data tell you? What other data might you need obtain to develop the analysis?” (s. 59).

⁷ Według numeracji ciągłej – rozdział 7.

⁸ A zatem rozdział 8. całej książki.

⁹ Czyli rozdział 9. – zgodnie z przyjętą numeracją.

(*strasznie wyglądać* ‘bardzo źle wyglądać; wyglądać przerażająco’ *vs.* *strasznie się podobać* ‘bardzo się podobać’).

Zamieszczone w trzeciej części książki przykłady wykorzystania korpusów w rozwiązywaniu określonych problemów językowych wydają się szczególnie ciekawe z badawczego punktu widzenia i skłaniają do sformułowania kilku refleksji. Pierwsza – ogólna – dotyczy, powiedzielibyśmy, natury korpusów: jakich wartości można w korpusach szukać? Refleksja druga odnosi się do gatunków tekstów, które korpusy gromadzą – czy są to gatunki literackie, czy opierające się na języku potocznym? Czy dostępne nam korpusy są w takim stopniu zrównoważone, by pozyskane z nich dane mogły stanowić reprezentatywną próbę? I w końcu refleksja chyba najważniejsza, bo dotycząca realnych możliwości wykorzystania korpusów językowych: w jaki sposób można przeprowadzić analizę otrzymanych z korpusu danych?

Reasumując, książka *Corpus Linguistic for Grammar*, opatrzona podtytułem *A Guide for Research*, w pełni i należycie spełnia funkcję przewodnika po metodzie badań korpusowych – wraz z ćwiczeniami. Napisana jest łatwym w odbiorze i zrozumiałym językiem. Zamieszczone w niej liczne wykresy, zestawienia, tabele i rysunki zachęcają do korzystania z narzędzi omówionych przez autorów. Zapowiedziany w podtytule i deklarowany przez autorów we wstępie praktyczny wymiar książki w zupełności zaspokaja oczekiwania odbiorcy. Ćwiczeń w *Corpus Linguistic for Grammar* jest bardzo dużo, a liczne odpowiedzi i komentarze do zadań dodatkowo motywują do głębszego rozważenia podejmowanych w poszczególnych rozdziałach kwestii. Dodatkowym atutem omawianej publikacji jest zamieszczony na końcu słowniczek trudnych pojęć (48 haseł), a także bogata literatura przedmiotu podawana po każdym rozdziale.

Lektura książki umożliwia dokładne zapoznanie się z kompetentnie przedstawionymi narzędziami korpusowymi i rozpoczęcie własnych badań z ich pomocą. Zajmując się językiem, warto znać te metody. Wykorzystanie korpusów zrewolucjonizowało bowiem sposoby eksploracji języka i gramatyki, a także dało możliwości poznawania różnych (mówionych lub pisanych) rodzajów i gatunków tekstu (kłótnia, publikacja biznesowa, tekst etyczny). Na podstawie wyników dostarczonych przez korpus możemy stwierdzić, czy istnieje ogólna tendencja do stosowania określonego wzorca – takiego, który stanowił założenie w badaniach, oraz sprawdzić, jak funkcjonuje ten wzorzec w tekstach zawartych w korpusie. Jak przekonują autorzy: korpus dostarcza danych językowych, jednak zawsze rolą badacza jest te dane odpowiednio zinterpretować¹⁰. Uważam, że warto z takich zasobów korzystać.

Literatura

- BIBER D., CONRAD S., REPPEN R., 2008: *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge.
 COLLINS L., 2019: *Corpus Linguistics for Online Communication*. London.

¹⁰ „This simply means that we, as researchers, will always have a role in interpreting the data a corpus can provide” – s. 15.

- GŁOWACKA E., 2011: *Jakość bibliotek cyfrowych – aspekty i kryteria oceny*. „E-mentor” nr 2 (39) [online: <http://www.e-mentor.edu.pl/arttykul/index/numer/39/id/828>; data dostępu: 30.11.2018].
- HANDFORD M., 2018: *Corpus Linguistics for Discourse Analysis*. London.
- LEWANDOWSKA-TOMASZCZYK B., red., 2005: *Podstawy językoznawstwa korpusowego*. Łódź.
- LÜDELING A., KYTÖ M., 2010: *Corpus Linguistics: An International Handbook*. Berlin.
- MCENERY T., HARDIE A., 2011: *Corpus Linguistics: Method, Theory and Practice*. Cambridge.
- NKJP: *Narodowy Korpus Języka Polskiego* [online: <http://nkjp.pl>; data dostępu: 30.11.2018].
- O’SULLIVAN J., 2019: *Corpus Linguistics and the Analysis for Sociolinguistic Change*. London.
- PRZEPÍÓRKOWSKI A., BAŃKO M., GÓRSKI R.L., LEWANDOWSKA-TOMASZCZYK B., red., 2012: *Narodowy Korpus Języka Polskiego*. Warszawa [online: http://www.nkjp.pl/settings/papers/NKJP_ksiazka.pdf; data dostępu: 30.11.2018].
- PRZEPÍÓRKOWSKI A., GÓRSKI R.L., LEWANDOWSKA-TOMASZCZYK B., ŁAZIŃSKI M., 2009: *Narodowy Korpus Języka Polskiego*. „Biuletyn Polskiego Towarzystwa Językoznawczego” LXV, s. 47–56.
- RÜHLEMANN Ch., 2018: *Corpus Linguistics for Pragmatics*. London.
- SINCLAIR J., 1991: *Corpus, Concordance, Collocation*. Oxford.
- TIMMIS I., 2015: *Corpus Linguistics for ELT*. London.
- WRAY A., 2005: *Formulaic Language and the Lexicon*. Cambridge.